



**UNIVERSIDAD
TORCUATO DI TELLA**

UNIVERSIDAD TORCUATO DI TELLA

DEPARTAMENTO DE ECONOMÍA

MAESTRÍA EN ECONOMÍA APLICADA

**Aprendizaje en matemática y programas mediados por
tecnología: revisión de la literatura empírica**

Alumno: Lucía Ethel Navarro, Legajo N° 18Y2177

Tutor: Hernán Ruffo

Fecha: 6 de junio de 2020

Una revisión de la evidencia sobre programas de aprendizaje mediado por tecnología en matemática

Resumen

Los programas de aprendizaje asistido por tecnología (CAL) han sido objeto de varios estudios. Este trabajo se propone revisar la evidencia existente sobre programas CAL orientados a la mejora en el desempeño en matemática en países en desarrollo. Tras revisar cinco casos relevantes para la literatura, los resultados encontrados enfatizan la importancia del diseño de una intervención – en qué medida es necesaria y factible en un contexto dado- y cómo se adapta a la necesidad de los alumnos y su heterogeneidad. En conclusión, la tecnología no deja de ser un medio de aprendizaje que, si no se ajusta a su contexto, no brindará soluciones mágicas.

Palabras Clave: educación, evaluación de impacto, tecnología, matemática.

Contenido

INTRODUCCIÓN	3
METODOLOGÍA DE LA REVISIÓN DE LITERATURA.....	4
REVISIÓN DE CASOS	6
RESULTADOS	21
DISCUSIÓN.....	26
CONCLUSIONES	27
REFERENCIAS.....	28

INTRODUCCIÓN

Son pocos los consensos que existen hoy en día sobre el uso de tecnología en ámbitos educativos, sobre los tipos de intervenciones tecnológicas en los que vale la pena invertir y en qué contextos (Escueta, Quan, Nickow & Oreopoulos, 2017). Esto se debe, en parte, a que hay pocas evaluaciones de impacto con evidencia robusta sobre el uso de programa de aprendizaje mediado por la tecnología. A su vez, las existentes presentan resultados que son, por lo menos, ambiguos. Por otro lado, la rapidez de lo tecnológico es enemiga de la lentitud con que los ámbitos educativos han sido capaces de adoptarlo. Esta rápida obsolescencia ha superado por mucho la capacidad de los investigadores de políticas para mantenerse al día con la evaluación (Escueta et al, 2017).

A pesar de lo anterior, el uso de la tecnología en el aula parece ser cada vez más prometedor cuando se trata de efectividad de la pedagogía (Mo, Zhang, Luo, Qu, Huang, Wang, Qiao, Boswell & Rozelle, 2014), en especial en países en desarrollo. De hecho, aunque escasa, la evidencia pareciera señalar resultados más alentadores en países en desarrollo que la que existe en países desarrollados (Escueta et al, 2017).

Algunos motivos por los que las tecnologías son prometedoras en contextos educativos podrían ser, en palabras de Glewwe & Muralidharan (2015), la posibilidad de replicar las intervenciones de forma costo-efectiva, de superar limitaciones de capacitación y formación docente, de enseñar contenidos avanzados o idiomas -cuya demanda es creciente, pero la oferta de docentes que los enseña es escasa-, de proveer instrucción en el hogar, de proveer modelos interactivos que motiven a los alumnos, de personalizar los planes de estudio y de reducir el tiempo en que los alumnos reciben devoluciones y comentarios con su trabajo.

Ahora bien, las intervenciones tecnológicas incluyen un amplio espectro de posibilidades. La tecnología no solo se utiliza para aplicaciones educativas, sino que también hay muchas herramientas asistidas por computadora que se utilizan para organizar y planificar el sistema educativo, así como herramientas para la evaluación e investigación de habilidades matemáticas y discapacidades (Räsänen, Laurillard, Käser & von Aster, 2019). A su vez, existen distintos tipos de categorizaciones respecto a las tecnologías, como, por ejemplo, 1) acceso a la tecnología, 2) aprendizaje asistido por computadora, 3) cursos en línea y 4) intervenciones conductuales (Escueta et al, 2017), o bien la de Cheung & Slavin (2013), que en su meta-análisis clasifican las intervenciones tecnológicas en 1) los programas CAL, 2) aplicaciones tecnológicas innovadoras, 3) sistemas de aprendizaje administrado por computadora y 4) modelos integrales, entre otras.

En este documento, el análisis se centrará en iniciativas de aprendizaje mediado o asistido por tecnología, usualmente conocidas como programas CAL en la literatura (por su nombre en inglés, *computer-assisted learning programs*). Estos pueden ser entendidos como un software o programa especialmente diseñado para el aprendizaje personalizado de habilidades específicas (Escueta et al, 2017). En general, estos programas proveen de instrucción y ejercicios en computadoras, ejercicios en entornos de bajos recursos docentes o de enseñanza; y que permiten proveer al niño de la enseñanza típicamente brindada por un docente (Mo et al, 2017), que en ocasiones es individualizada y adaptable a su nivel de conocimientos. No siempre tienen, sin embargo, las mismas características: algunos permiten el aprendizaje adaptativo, otros implican el trabajo colaborativo entre alumnos, otros incluyen elementos lúdicos, entre otras.

Los programas de aprendizaje asistido por tecnología (programas “CAL” de ahora en más) han sido objeto de varios estudios. Este estudio se propone revisar la evidencia existente sobre programas CAL orientados a la mejora en el desempeño en matemática en países en desarrollo¹.

A continuación, se presenta primero la metodología empleada para la revisión de literatura, se revisan los cinco casos seleccionados, se analizan y comparan los resultados obtenidos y finalmente se presenta una discusión sobre dichos resultados.

METODOLOGÍA DE LA REVISIÓN DE LITERATURA

Este estudio revisa la evidencia existente sobre programas CAL orientados a la mejora en el desempeño en matemática de alumnos en países en desarrollo.

La evidencia disponible en países en desarrollo sobre los programas CAL varía tanto en las direcciones como en sus magnitudes, aun cuando los efectos provienen de evaluaciones aleatorizadas rigurosas (Glewwe & Muralidharan, 2015). Esto podría señalar la importancia del contexto y del diseño de estos programas al analizar la efectividad de cada intervención (Glewwe & Muralidharan, 2015). Por ese motivo, a diferencia de otras revisiones, este estudio busca comparar las características de intervenciones relevantes y rigurosas llevadas adelante en países en desarrollo, para identificar si hay y cuáles son los patrones que podrían explicar efectos positivos en el desempeño en matemática. ¿Por qué matemática? Por un lado, la evidencia señala que es en esta área curricular donde los programas CAL podrían tener mejores resultados (Escueta et al, 2017; Kulik, 2003; Murphy, Penuel, Means, Korbak, Whaley & Allen, 2002). Por otro lado, la matemática es cada vez más requerida en la economía digital y, por lo tanto, más necesaria que nunca.

La restricción de que el análisis incluya únicamente países en desarrollo responde a dos necesidades de la literatura existente. En primer lugar, la mayoría de los experimentos aleatorizados fueron realizados en países desarrollados, cuyos contextos son hoy muy lejanos a los de países en vías de desarrollo. Por este motivo, las intervenciones evaluadas en la literatura difieren demasiado para compararlas entre sí (Escueta et al, 2017). En segundo lugar, la poca evidencia existente señala que las dificultades socio-económicas y características educativas propias de los países en vías de desarrollo podrían favorecer la efectividad de los programas CAL en el aprendizaje (Escueta et al, 2017). Las grandes desigualdades en los primeros, junto con la

¹ Con el foco puesto en los programas de aprendizaje asistido por tecnología, esta revisión excluye las intervenciones centradas en el hardware. Éstas, usualmente desarrolladas en la década del 2000, refieren a políticas enfocadas en proveer infraestructura tecnológica, es decir, en la distribución de equipos como notebooks, netbooks, computadoras de escritorios en la forma de “laboratorios”, tablets, entre otros. Si bien en muchos casos contenían softwares para el aprendizaje, el énfasis estaba puesto en la infraestructura, y no así en el programa que proveían en cuestión. A su vez, existe evidencia y por tanto consenso de que la mera introducción de la tecnología no es suficiente para mejorar los resultados académicos (Escueta et al, 2017; Barrera-Osorio & Linden, 2009). Un ejemplo fue el programa de uno-a-uno conocido como One Laptop Per Child en Perú (Cristia, Ibararán, Cueto, Santiago & Severin, 2012), cuya evaluación demuestra que la distribución de dispositivos tecnológicos no tuvo efectos en el aprendizaje de matemática ni de lengua.

a veces insuficiente oferta de docentes, la baja motivación y/o capacitación docente y la escasez de recursos en general podrían ser ventanas de oportunidad para el uso de tecnologías².

La revisión incluye intervenciones en nivel primario y medio de la enseñanza. Sin embargo, hay una clara mayoría de intervenciones en nivel primario. Esto podría deberse a que, durante la década del 2.000, hubo claras políticas de expansión del acceso a nivel primario, derivando en un incremento de la matrícula. Si bien aún hay numerosos desafíos en este nivel, el hecho de que no haya o de que haya menos barreras de acceso, ha permitido que las evaluaciones aleatorizadas en este nivel se llevaran adelante en poblaciones vulnerables que, en otro tiempo, no hubieran podido estar en la escuela. En cambio, el nivel secundario aún presenta grandes complicaciones en términos de acceso y finalización; lo cual pudo haber funcionado como desincentivo a realizar este tipo de evaluaciones.

A su vez, esta revisión de literatura incluye únicamente evaluaciones aleatorizadas que analizaron los efectos de programas CAL en los desempeños en matemática. En algunos casos, también se analizaron otras áreas (por ejemplo, lengua o contenidos cognitivos), pero los resultados en éstas no fueron tenidas en cuenta en este análisis. Del mismo modo, aquellos estudios que evaluaran contenidos no matemáticos, fueron dejados de lado para el análisis.

Al ser pocas las evaluaciones rigurosas existentes en países en desarrollos sobre el tema, el análisis presentado en esta revisión abarca gran parte de los estudios aleatorizados llevados adelante en países en desarrollo desde el 2005³. Dada la rápida obsolescencia de la tecnología, este límite temporal busca que el análisis incluya tecnologías que, aunque viejas, puedan compararse a pesar de los años transcurridos entre ellos. Es decir, pretende evitar la comparación de intervenciones tecnológicas de los '90 con otras, por ejemplo, del 2015. A su vez, este rango permite abarcar gran parte de la literatura relevante sobre programas CAL.

Los estudios incluidos en esta revisión fueron realizados en su mayoría en India, donde se han llevado adelante el mayor número de investigaciones, aunque también se incluyen evaluaciones aleatorizadas realizadas en China, Chile y Ecuador. Cada investigación presenta características propias en su diseño e implementación, las cuales se detallan para una mejor comprensión sobre cómo se implementa la tecnología en el ámbito educativo en cada caso. A lo largo del análisis, el énfasis estará puesto en los siguientes aspectos de cada intervención: a. La intervención y estrategia de identificación, b. diseño del software y contenido, c. Target de la intervención, d. relación software-aula, e. evaluación del desempeño y datos recolectados, f. Resultados.

² Es preciso aclarar que el aglomerado de “países en desarrollo”, si bien útil para el análisis, también contiene una gran heterogeneidad a su interior; pues abarca países con enormes diferencias en términos de sus sistemas educativos, entre otras. Por tanto, aun limitando la revisión a éstos, se mantienen diferencias en la implementación de los programas CAL en cada país.

³ Lejos de ser un análisis exhaustivo de la literatura, aquí nos enfocamos en aquellos estudios que, de gran magnitud, con metodologías rigurosas y diseños confiables, permitan extraer conclusiones sobre el uso de programas CAL en tecnología. Dada la evidencia revisada, se encuentra que son pocos los estudios que cumplen con las restricciones impuestas en esta revisión, por lo que solo se excluyó una investigación llevada adelante en china (Lai et al, 2015) cuyas características y conclusiones eran similares a las de casos analizados (Mo et al, 2014).

La sección de intervención y estrategia de identificación de cada caso describe, en breves palabras, en qué consiste la intervención y la forma de seleccionar a los participantes.

El diseño y contenido del software es un punto central del análisis. Un programa puede contemplar diversos mecanismos de interacción del alumno con el contenido y el programa. Así, por ejemplo, habrá programas que se caractericen por la adaptabilidad de estas plataformas a la velocidad y nivel de aprendizaje del niño, mientras que otros se centrarán en proveer contenidos idénticos para todos los alumnos y reforzar lo visto en clase.

Por otro lado, el *target* de la intervención es una sección donde se distingue claramente a los destinatarios de las intervenciones; con el objetivo de comprender mejor cómo el programa busca satisfacer – o no- sus necesidades en particular.

Asimismo, la relación de estos programas con los contenidos y recursos existentes parece ser un factor de éxito (Glewwe & Muralidharan, 2015). Con sistemas educativos que han enfrentado pocos cambios a lo largo de la historia (Glewwe & Muralidharan, 2015), las intervenciones adecuadas al sistema en el que se insertan podrían aprovechar y potenciar las fortalezas de cada sistema educativo.

Finalmente, la evaluación de los contenidos resulta sumamente importante. Cómo se evalúan los contenidos matemáticos y qué contenidos se evalúan podría explicar por lo menos parcialmente la diferencia entre los resultados de los casos. Así, por ejemplo, una evaluación de “amplio espectro” de los contenidos al inicio y final de una intervención (pre y post) permitirá captar el desempeño de alumnos que están por debajo del nivel de su grado; mientras que evaluaciones con un nivel mínimo demasiado alto, no podrán identificar cambios (Banerjee et al, 2017; Muradliharan, Singh & Ganimian, 2017).

REVISIÓN DE CASOS

A continuación, se analizan 5 evaluaciones aleatorizadas llevadas adelante en países en desarrollo:

2.1. Remedying education: evidence from two randomized experiments in India (Banerjee, Cole, Duflo & Linden, 2007).

Los autores analizan en un mismo estudio dos intervenciones llevadas adelante en zonas urbanas de India. La primera, el programa *Balshaki*⁴, analiza los efectos de un programa de compensación de contenidos (*remedial education*). La segunda intervención evalúa los efectos de un programa de aprendizaje asistido por computadora en niños de 4° grado, sin discriminar

⁴ El programa Balshaki consistió en brindarle instrucción de matemática a alumnos de bajos rendimientos de 3° y 4° grado. La instrucción era brindada por mujeres locales poco calificadas (secundario terminado), entrenadas específicamente para este programa (denominadas balshakis, de allí la forma de referirse al programa). De este modo, durante el horario escolar se retiraban a los alumnos de bajos rendimientos por dos horas, para recibir en grupos de 15-20 alumnos instrucción por las balshakis, quienes trabajaban sobre contenidos básicos de 1° y 2° grado que no habían sido comprendidos del todo. Se implementó durante 2001 y 2002 (Banerjee et al, 2017).

por bajo, alto o medio desempeño escolar. A los efectos de esta revisión, sólo se tomará lo relacionado con el segundo análisis⁵.

a. La intervención y su estrategia de identificación

El estudio se llevó adelante en la ciudad de Valdodara, India. El programa CAL se implementó durante dos años, en dos cohortes distintas de alumnos de 4° grado: 2002-2003 y 2003-2004. La implementación del proyecto fue realizada por Prathram, una fundación sin fines de lucro local.

En ambos años, la actividad para el grupo tratamiento fue la misma: los alumnos utilizaron las computadoras dos horas la semana, una de ellas durante el horario de clase y otra al principio o al final de la jornada escolar. En cada sesión, dos alumnos compartían una computadora. Se utilizaron 4 computadoras en cada escuela pública, las cuales ya estaban disponibles gracias a una política pública llevada adelante en el 2000. La intervención duró todo el año escolar.

El grupo de control tenía acceso a las computadoras para usarlas libremente, aunque según los autores no fueron utilizadas.

La muestra estuvo compuesta por las escuelas que participaron del programa balshaki. La mitad de las escuelas del grupo tratamiento y la mitad del grupo control conformaron el grupo tratamiento del programa CAL. Las restantes fueron asignadas al grupo control. Esta asignación excluyó a las escuelas que no disponían de computadoras. De esta forma, en el año 2002-2003 la muestra analizada se compuso de los siguientes grupos, asignados por *clusters* -es decir, por escuelas-:

- Grupo A1B1 (tratamiento – año 1): 2.850 estudiantes de 4° grado de 55 escuelas municipales
- Grupo B1A1 (control – año 1): son 3.095 estudiantes de 4° grado de 56 escuelas municipales

La asignación fue aleatorizada y se estratificó por las variables control/tratamiento en el programa balshaki, género, idioma de enseñanza en la escuela y desempeño en matemática según las evaluaciones del año anterior.

En el año 2003-2004, se mantuvo el programa, pero se invirtieron los grupos tratamiento y control, para los nuevos 4° grados de las escuelas de participantes. Es decir, los grupos se conformaron establecidos de la siguiente forma:

- Grupo A1B1 (control -año 2): 2814 estudiantes de 4° grado de 55 escuelas municipales
- Grupo B1A1 (tratamiento – año 2): 3.131 estudiantes de 4° grado de 56 escuelas municipales.

Como el programa Balshaki se llevó adelante en dos cohortes, en los años 2001-2002 y 2002-2003, los estudiantes participantes de la primera cohorte del programa CAL participaron también de la segunda cohorte del programa balshaki. Sin embargo, debido al diseño muestral, sólo los alumnos con bajos rendimientos participaron de este último, y conformaron una parte

⁵ Es posible separar el análisis de las dos intervenciones gracias al diseño experimental utilizado por los autores, como se verá más adelante.

similar tanto del grupo tratamiento y control. Durante 2003-2004, la única intervención activa fue el estudio del programa CAL⁶.

La muestra estuvo balanceada debido a la estratificación. Salvo en los resultados del *pretest* del del segundo año del grupo de tratamiento -superior en 0.13 DS al desempeño del grupo control-, no hubo diferencias estadísticamente significativas en los exámenes de inicio de año.

b. Diseño del software y contenido.

El software consistió en juegos de resolución de problemas matemáticos cuyo nivel de dificultad se adaptaba a la capacidad de los alumnos para resolverlos. El mismo fue especialmente diseñado para reforzar las competencias básicas del currículum⁷.

Los encuentros fueron mediados por instructores⁸, cuyo rol era facilitar las sesiones de juego, incentivar a los alumnos a utilizar los juegos de las computadoras y asistirlos en el caso de que no comprendieran la consigna de algún juego - pero sin proveer instrucción de matemática-.

c. *Target* de la intervención

La intervención estaba dirigida a alumnos de 4º grado de nivel primario, de escuelas urbanas y municipales de la ciudad de Vadodara, India. No hay una descripción del nivel promedio de estas escuelas en relación al municipio y/o país.

d. Relación software-aula

El uso del software se dio tanto dentro como fuera del horario de clase, aunque no hubo más allá del espacio temporal otra vinculación con el aula durante la implementación del programa.

Con un software cuyo contenido estaba basado en competencias básicas del currículum nacional, los docentes no intervenían durante la intervención, ya que para eso fueron capacitados los instructores.

e. Evaluación del desempeño y datos recolectados

⁶ El Programa Balshaki se implementó durante 2001 y 2002; y participaron 98 escuelas municipales del experimento. El primer año, los 3º grados de la mitad de estas escuelas fueron asignados al grupo tratamiento; mientras que los 3º grados del resto de las escuelas fueron asignados al grupo control. A su vez, de estas últimas – donde el 3º grado era control-, los alumnos de 4º grado se asignaron al grupo tratamiento. De este modo, las escuelas de tratamiento en 3º grado eran las de control del tratamiento en 4º grado y viceversa. En 2002, se invirtió la asignación: las escuelas que habían implementado el programa Balshaki en 4º grado, ahora lo implementaban en 3º; mientras que las que lo habían implementado en 3º grado, ahora lo implementaban en 4º. Nuevamente, esto generaba grupos de control inversos, así como un único grupo que recibió el tratamiento durante dos años: 3º grado de las escuelas asignadas a tratamiento en el año 1, que luego fueron tratamiento en 4º grado en el año dos -es decir, misma cohorte de alumnos- (Banerjee et al, 2017).

⁷ El primer año, se utilizó un software desarrollado por Pratham, mientras que en el segundo año Pratham se alió con una empresa de software local para hacer modificaciones que permitieran un seguimiento más preciso del currículum escolar (Banerjee et al, 2017).

⁸ Pratham fue la encargada de la selección y capacitación en uso de tecnología de un equipo de instructores (Banerjee et al, 2017).

El desempeño escolar fue medido a través de evaluaciones anuales que se realizaron antes y al final de cada año escolar. Dado el diseño de intervención, se obtuvieron mediciones durante dos años. Aunque el tratamiento se asignó a distintas cohortes, los autores obtuvieron mediciones de corto y “largo” plazo: con el grupo del primer año, pudieron medir la persistencia del efecto al tomar evaluaciones un año después de finalizada la intervención (2003-2004); aunque fuera con una muestra más chica. Los contenidos de estos instrumentos abarcaron los de 1º a 4º grado, y fueron administrados por el equipo de investigación.

f. Resultados

Para realizar el análisis, los autores estiman la intención al tratamiento (o *IIT* por las siglas en inglés de *intention-to-treat*) mediante una regresión lineal de los resultados en el examen final, controlando por el examen inicial. En el corto plazo, el programa tuvo efectos de 0.35 desvíos estándar (DS de ahora en adelante) en comparación con el grupo control durante el primer año, mientras que en el segundo año el grupo tratamiento obtuvo resultados 0.47 DS más altos que el grupo control, una vez corregido por las diferencias en el examen inicial. Los errores estándares fueron de 0.076 y 0.068 para el año 1 y 2 respectivamente, los cuales fueron calculados a nivel escuela, al igual que en todo el análisis. El intervalo de confianza al 95% del primer año se encuentra entre el rango {0.20104, 0.49896}, el del segundo año entre {0.33672, 0.60328}⁹.

Al ver la distribución de estos resultados según el desempeño la evaluación inicial – sólo de la cohorte 2002-2003 -, es posible ver que el programa benefició más a los alumnos con rendimientos más bajos, luego a los de distribución media y finalmente a los mejores de la clase. Así, los alumnos con desempeños más bajos obtuvieron mejoras de 0.425 DS (con error estándar 0.106), aquellos con desempeños medios mejoras de 0.316 DS (0.081 de error estándar) y los alumnos con mejor rendimiento mejoraron sólo 0.266 DS (0.073 de error estándar).

Sobre el nivel de *attrition*, es bajo y se encuentra distribuido de manera homogénea entre tratamiento y control (3,8% y 3,4%, para tratamiento y control respectivamente en el año 1, y 7,9% y 6,9% en año dos). A su vez, las características de los desertores son similares entre grupos según los autores.

En cuanto al largo plazo, los autores analizan el efecto del programa CAL un año después de que la intervención finalizó. Para ello, estiman la diferencia entre exámenes tomados al final del 2004 (evaluación final, dos años después de la inicial) con los del inicio de año del 2002 (evaluación de línea de base). Los efectos son significativos y con un valor mucho más bajo que el de corto plazo: 0.092 DS de efecto promedio, con un error estándar de 0.045. Asimismo, cuando se analiza la persistencia en el desempeño de los alumnos en el tercil con peores desempeños, encuentran un resultado levemente mejor para ellos, con un valor de 0.107 DS (0.046 de error estándar, a nivel escuela). Los intervalos de confianza al 95% se encuentran entre los rangos {0.0038, 0.1802} y {0.01684, 0.19716} respectivamente.

⁹ Como el estudio revisado no menciona ni el nivel de significatividad ni el intervalo de confianza para cada estimación, se construyeron los intervalos de confianza al 95%, empleando la fórmula *coeficiente coeficiente* $\pm 1.96 * error\ estándar$. El mismo procedimiento se aplicó para el resto de los resultados presentados de este estudio.

2.2. Disrupting education? Experimental evidence on technology-aided instruction in India (Muradliharan, Singh & Ganimian, 2017).

Los autores analizan el efecto de un programa extraescolar de *blended learning*¹⁰, que consistió en el uso del software Mindspark y en la instrucción grupal suplementaria.

a. La intervención y su estrategia de identificación

La evaluación de impacto se llevó adelante en tres centros (*Mindspark centers*) de Nueva Delhi, espacios especialmente destinados al uso del programa de forma suplementaria a la educación formal, durante 2015-2016.

La intervención consistió en la asistencia de alumnos de nivel medio a estos centros seis veces a la semana, fuera de horario escolar, durante cuatro meses y medio. Las sesiones de entrenamiento duraron 90 minutos diarios, donde 45 minutos se destinaban al uso Mindspark y los otros 45 minutos a una instrucción grupal, de a grupos de 12-15 estudiantes. Los contenidos de estas sesiones estaban distribuidos por día: dos días matemática, dos días hindi, un día inglés y un día opcional si los alumnos lo querían. Por tanto, a efectos de analizar los desempeños en matemática, podría contemplarse que el uso del programa CAL implicó sesiones de 90 minutos, dos veces por semana. Los encuentros grupales, llevados adelante por instructores locales, consistían ayudar a los alumnos a resolver la tarea y a repasar los conceptos básicos escolar, sin customización alguna.

Para conformar la muestra, se invitó a los alumnos de los grados 4° a 9° de cinco escuelas públicas¹¹ cercanas a los centros a inscribirse para participar de un sorteo para participar del uso del programa CAL. La mitad de los inscriptos fue asignada aleatoriamente a los centros Mindspark, es decir, del grupo tratamiento; mientras que la otra mitad estarían en el grupo que no podría asistir, es decir, del grupo control. Los ganadores del sorteo recibieron *vouchers* para acceder al programa de forma gratuita, mientras que los que no fueron elegidos recibieron acceso gratuito a los centros una vez finalizada la intervención – si y solo si habían participado de la evaluación final-.

En total, se presentaron 619 estudiantes de los grados 6° a 9° en su mayoría. La asignación aleatorizada fue estratificada por centro Mindspark, priorizando las preferencias en días y horarios que cada alumno declaró, de forma tal de asegurar la mayor asistencia posible luego. En consecuencia, 305 alumnos fueron asignados al grupo control y 314 alumnos asignados al grupo tratamiento¹². No hubo diferencias significativas en variables descriptivas (edad, género, nivel socio-económico) entre los grupos.

b. Diseño del software y contenido.

¹⁰ Es decir, que combina la enseñanza tradicional con enseñanza no tradicional que, en este caso, es con tecnología.

¹¹ De las cuales tres eran escuelas sólo para mujeres, mientras que dos eran escuelas masculinas únicamente (Muradliharan, Singh & Ganimian, 2017).

¹² La muestra, autoseleccionada por el mecanismo de inscripción al sorteo, presentó niveles levemente superiores que el resto de los alumnos de las cinco escuelas involucradas en el estudio, en los exámenes finales del año anterior -es decir, los de 2014-2015- (Muradliharan, Singh & Ganimian, 2017).

El software fue desarrollado por una empresa de educación local, con un desarrollo que llevaba más de diez años al momento de la evaluación. A diferencia de otros programas, son varias las características destacables de Mindspark: (1) adaptación de las tareas y complejidad del juego según el nivel y la tasa de progreso de cada alumno, (2) a partir de los datos de juego, identifica patrones de errores de los alumnos y enfatiza las tareas en estos contenidos, (3) podía usarse con o sin internet, (4) estaba basado en investigación en pedagogía efectiva para ambientes socio-económicos vulnerables, y (5) podía utilizarse tanto en computadoras, tablets o celulares.

Su contenido se diseñó en base al currículo local, aunque la complejidad del juego estaba determinada por el nivel de desempeño del alumno -y no por el currículum del grado al que pertenecía-.

c. *Target* de la intervención

La intervención involucró a alumnos de bajo nivel socio-económico, que no hubieran podido cubrir el costo de asistir a un centro Mindspark de otra forma. A su vez, estos alumnos debían estar entre los grados de 6° a 9° grado del nivel medio.

d. Relación software-aula

El programa tuvo lugar fuera del horario escolar. La relación con el aula estuvo dada, por un lado, por el *curriculum-based software* y, por otro y de forma más cercana, por la instrucción grupal donde se realizaba la tarea de la escuela y se ayudaba a prepararse para exámenes.

e. Evaluación del desempeño y datos recolectados

Se tomaron exámenes en papel y lápiz en las materias de matemática e hindú, a fines del año anterior de empezar la intervención a todos los alumnos de las escuelas involucradas en el estudio, y al final del año escolar en el que se llevó la intervención. Fueron evaluaciones especialmente diseñadas para este proyecto, cuyo nivel cubrió un amplio rango de niveles de desempeño de los alumnos.

Por otro lado, utilizaron datos administrativos de las evaluaciones escolares de finales del año anterior al estudio de todos los alumnos de las escuelas participantes.

Por último, a través de los datos del software se obtuvieron los datos de uso del programa, del contenido presentado, de las respuestas a cada pregunta y del nivel de grado estimado del logro del estudiante en cada momento. Esta información se utilizó para analizar las brechas en desempeño y heterogeneidad de los alumnos.

f. Resultados

Los autores utilizan una regresión lineal con efectos fijos – de cada estrato- robusta para la estimación del *intention-to-treat (ITT)*. El programa obtuvo mejoras de $0.36\sigma^{13}$ en matemática en comparación con los alumnos del grupo control, luego de 4.5 meses de la intervención. El error estándar robusto fue de 0.063 a nivel escuela, y el nivel de confianza del 99%.

¹³ Los puntajes de los diversos grados se transformaron con la escala de la Teoría de Respuesta al Ítem y fueron estandarizados. Por tanto, los resultados están normalizados (Muradliharan, Singh & Ganimian, 2017).

También especifican la regresión con controles por edad, género y nivel socio-económico, sin encontrar diferencias significativas. Incluso con análisis no paramétricos de los datos, encuentran resultados similares a los de ITT. Además, los autores estiman el efecto del programa con la asistencia como variable instrumental, con el objetivo de incluir el nivel de asistencia efectiva a los centros Mindspark en el análisis – en vez de la asistencia promedio, ya que la asistencia a los centros de los alumnos de tratamiento fue en promedio de 58%. -. Asumiendo efectos de tratamiento constantes y una respuesta lineal por día de tratamiento, los autores estiman un efecto promedio en los tratados (*average treatment effect, o ATE* por sus siglas en inglés) que en 90 días de la intervención podrían resultar en mejoras en matemática de 0.59σ con tan solo un 80% de asistencia, nuevamente con un nivel de confianza del 99% y con un error estándar de 0.062 a nivel escuela.

A su vez, si bien los autores no encuentran diferencias al analizar los resultados por tercios de niveles de desempeño, sí hallan que los alumnos del grupo control que componen el tercio con desempeños más bajos de cada grado no presentan mejoras durante el transcurso del año escolar. De hecho, esto parece aún más dramático si se tiene en cuenta que, según los datos recolectados en el estudio, todos los alumnos poseían un nivel en matemática que es en promedio 2.5 grados más bajo del que deberían tener para su grado. Esta situación podría exagerar los resultados de la intervención, puesto que la comparación se da entre el tratamiento y la ausencia de intervención, la cual presenta resultados muy bajos.

Por otro lado, al analizar los datos del software, los autores señalan que los efectos positivos se encontraron en las preguntas que estaban por debajo del nivel del grado que le correspondía, mientras que no hubo efecto alguno en las preguntas del nivel correspondiente.

Respecto al nivel de *attrition*, solo el 86% de los alumnos asistieron al examen final, correspondiendo un 84% al grupo tratamiento y 88% al grupo control. Como la diferencia es significativa al 10%, utilizan un estimador de probabilidad inversa para analizar el efecto de la intervención. Los estimados IIT permanecen similares al utilizar un ponderador de probabilidad inversa.

Por último, es importante destacar que la intervención refleja no sólo el uso del software, sino también la instrucción grupal. Sin embargo, los autores atribuyen el efecto principalmente al uso del programa¹⁴.

2.3. ¿Funciona la gamificación en la educación? (Araya, Ortiz, Bottan & Cristia, 2019).

Se trata de una evaluación de impacto de un programa CAL, utilizado en horario escolar con niños de bajos rendimientos de escuelas primarias de Chile.

a. La intervención y estrategia de identificación

¹⁴ Los estudiantes fueron asignados a los centros Mindspark en lotes que a menudo incluían estudiantes matriculados en los grados 1-5 en el mismo momento. Esto se debe a que la consideración principal en la asignación fue la conveniencia de los estudiantes y los padres. Sin embargo, en términos de enseñanza en el aula, esto tiene muy poco sentido para instrucción grupal en palabras de los autores (Muradlihan, Singh & Ganimian, 2017).

La intervención consiste en una evaluación de impacto de un programa llamado Conecta Ideas, que busca mejorar el aprendizaje de alumnos de cuarto grado con bajo rendimiento de escuelas públicas de Santiago, Chile.

El proyecto consistió en el uso de una plataforma online con ejercicios en sesiones semanales de 90 minutos, durante el horario escolar regular y por 7 meses. Las sesiones se llevaban adelante en un laboratorio de computación, una de ellas en el horario de la clase de matemática, mientras que la otra representaba tiempo adicional de clase de matemática – en detrimento de otra clase, pues aún tenían lugar en horario escolar-. Así, los alumnos tenían dos sesiones por semana.

Se llevó adelante en 24 escuelas públicas, cuyos alumnos provenían de un contexto vulnerable y tenían retrasos en el aprendizaje. Para la selección de las escuelas, primero se identificaron 22 comunas con escuelas que cumplieran los requisitos mencionados, que debían estar dentro de las dos categorías más bajas de estatus socio-económico del Ministerio de Educación y que tuvieran 2 o más secciones de 4° grado de nivel primario. Luego, se identificaron comunas con directores comunales y directores de escuelas interesados, se realizaron sesiones informativas, se programó una visita y se visitaron escuelas para analizar si cumplían con los requisitos técnicos¹⁵. El proceso, al parecer inconcluso, terminó con 24 escuelas participantes de la evaluación, las cuales eran similares a una muestra de escuelas de status socio-económico bajo en Santiago; a excepción por la matrícula y los desempeños promedio de cada escuela¹⁶.

La asignación de secciones de 4° grado a tratamiento o control se realizó aleatoriamente, con *clusters* por sección. Así, se determinaron 24 secciones del grupo tratamiento, que usarían el programa CAL, y 24 secciones del grupo control, las cuales solo tendrían enseñanza tradicional.

Los grupos resultantes difirieron significativamente en dos variables: resultados en la base línea de matemática, donde el grupo tratamiento obtuvo notas 0.08 DS menores que el grupo de control; y en educación de la madre.

b. Diseño del software y contenido.

El programa fue diseñado por la Universidad de Chile con el objetivo de generar grandes aumentos en el aprendizaje de matemática de alumnos de bajo nivel socio-económico de Chile.

El programa consistía en sesiones donde se les asignaban a los alumnos el mismo conjunto de 20 a 30 ejercicios, los cuales estaban relacionados con el currículum y lo que vieron en clases

¹⁵ El proceso de selección comenzó a finales de enero de 2017 con la identificación de 22 comunas que tenían escuelas que satisfacían los criterios descritos más arriba. Los directores de las comunas fueron contactados primero por correo electrónico y luego por teléfono. El equipo de ConectaIdeas luego visitó a los 11 directores que contestaron y manifestaron su interés. Después, se llevaron a cabo sesiones informativas con los directores de las comunas y los directores de escuela en nueve comunas. En el paso final, el equipo llevó a cabo visitas técnicas en las escuelas para verificar los requisitos técnicos. Las visitas técnicas estaban programadas en 31 escuelas en 6 comunas. Sin embargo, después de visitar cuatro comunas, el equipo confirmó 24 escuelas que cumplían con los requisitos arriba mencionados y, por lo tanto, se puso fin al proceso de selección (Araya, Arias Ortiz, Bottan & Cristia, 2019).

¹⁶ La matriculación en las escuelas del estudio fue mayor (debido al requisito de las dos secciones) y sus alumnos tenían resultados aún más bajos en matemática y lenguaje (Araya, Arias Ortiz, Bottan & Cristia, 2019).

esa semana en cuestión. La plataforma enviaba retroalimentación automática respecto a si sus respuestas eran correctas o no al finalizar.

La plataforma se diseñó con elementos propios de la *gamificación*, que incrementaran la motivación por aprender matemática. Estas características fueron (1) resultados alcanzados visibles, para que los alumnos puedan seguir sus avances, y comparación con el resultado promedio de su sección; (2) mensajes motivacionales en el medio del juego, transmitiendo la idea de que la inteligencia es maleable y requiere de su esfuerzo mejorar; (3) motivación grupal a través de competencias entre secciones – medidas por el número promedio de ejercicios que realizaban por semana-; (4) a través de la colaboración en torneos simultáneos donde se compite con otras secciones para ver quién resuelve más ejercicios.

c. *Target* de la intervención

Alumnos de 4° grado con bajos rendimientos de escuelas primarias de Santiago, Chile.

d. Relación software-aula

La plataforma se utilizó en horario escolar, tanto dentro de la clase de matemática como en horas adicionales de clase (que se les restaban a otras materias). Se integra principalmente de dos formas al aula: (1) el docente, junto un coordinador capacitado sobre el programa, era quien asignaba las tareas a los alumnos; y (2) el software le proveyó a los docentes un panel de control con información en tiempo real sobre el número de preguntas contestadas y el número de preguntas correctas de cada alumno; y con información en general sobre el desempeño para identificar a aquellos que necesitan más apoyo y las preguntas en las que en general hubo un desempeño bajo. Además, el sistema genera informes de retroalimentación para los coordinadores del laboratorio, los docentes y los directores de escuela.

e. Evaluación del desempeño y datos recolectados

Utilizaron datos provenientes del examen nacional estandarizado en Chile, de evaluaciones a los participantes, de encuestas y de la plataforma ConectaIdeas. Respecto al examen nacional, se trata de una evaluación en papel implementada anualmente en todas las escuelas del país para monitorear el aprendizaje en matemática y lengua¹⁷. A su vez, esta información se complementó con evaluaciones al inicio, medio (4 meses tras iniciada la intervención) y fin del proyecto. Por otro lado, se recogieron datos a través de encuestas sobre las percepciones de los alumnos¹⁸, los cuales se complementaron con los de las encuestas de los exámenes nacionales estandarizados. Finalmente, el *engagement* de los alumnos del grupo tratamiento con la intervención fue analizado a través de los datos provistos por la plataforma ConectaIdeas.

¹⁷ Estas evaluaciones conllevan, a su vez, sus riesgos en términos de moral-hazard pues “docentes y los directores porque están vinculadas a incentivos monetarios” (Araya, Arias Ortiz, Bottan & Cristia, 2019).

¹⁸ Se levantaron datos de los alumnos respecto del autoconcepto en matemática, la motivación intrínseca en matemática, la preferencia por tener clases de matemática en el laboratorio de computación (en vez de en el aula regular), tener una mentalidad de crecimiento y la preferencia por el trabajo en equipo (Araya, Arias Ortiz, Bottan & Cristia, 2019).

f. Resultados

Los autores utilizaron dos especificaciones para estimar el *intention-to-treat (ITT)*: una regresión con efectos fijos y esa misma regresión utilizando la variable de resultado en la línea de base como control. A su vez, estas especificaciones se aplicaron con los datos de los exámenes nacionales, por un lado; y con los datos de los exámenes tomados en el marco del proyecto por otro.

Al analizar los exámenes nacionales, con la primera especificación se obtuvo que los alumnos del grupo tratamiento obtienen efectos positivos de 0.22 DS en comparación con el grupo control, con un error estándar de 0.05 a nivel sección escolar y un nivel de significancia del 1%. Con la segunda, el efecto es de 0.27 DS, con un error estándar de 0.04, significativo al 1% nuevamente. Al analizar la heterogeneidad de estos resultados, no se encuentran diferencias significativas por género, educación de la madre y resultados académicos en la línea de base.

Al emplear los exámenes del proyecto, los resultados son más bajos¹⁹: 0,09 DS no significativos para la primera especificación y 0,13 DS para la segunda con un nivel de significatividad del 1% y 0.05 de error estándar a nivel sección.

Es preciso aclarar, a su vez, los datos de registro del software muestran que la plataforma se utilizó en promedio por 27 horas, es decir, 43 días durante 39 minutos cada vez. Además, en el 98% de los casos en el marco de la escuela y su horario de clase.

Las tasas de *attrition* para estimar los efectos en los resultados de matemática fueron del 10% para el grupo de tratamiento y del 8% para el grupo de control.

2. 4. Complement or Substitute? The Effect of Technology on Student Achievement in India (Linden, 2008)

Se trata de una evaluación de impacto tanto sobre de un programa de aprendizaje computarizado como de la efectividad de la forma de implementarlo.

a. La intervención y su estrategia de identificación

La intervención tuvo lugar en Gujarat, India; y se hizo en el marco del programa Gyan Shala²⁰ de una ONG local. A su vez, el programa emplea como maestras a mujeres que no cumplirían los

¹⁹ Estos resultados son mayores cuando se realiza la diferencia entre la línea de base y los exámenes intermedios. Los autores argumentan que esto podría haber sucedido por los contenidos que los exámenes del proyecto evaluaban, los cuales no tenían su correlato con lo que la plataforma trabajaba.

²⁰ Este programa se encuentra dirigido a alumnos de 1° a 3° grado cuyas familias son de bajos recursos. La enseñanza se rige por un currículum especialmente diseñado por la ONG para seguir los contenidos básicos, que es implementado de forma rigurosa. Además, con la misma duración del ciclo escolar que el resto de las escuelas indias, su duración diaria es menor, pues los niños van solo 3 horas escolares por día. Por último, si bien este modelo optimiza los costos aun brindando materiales y libros para su uso en la escuela, Gyan Shala cubre la mayor parte del costo para que los niños asistan a la escuela, cobrándoles tan solo una pequeña parte o, en el caso de familias extremadamente pobres, sin costo alguno. De esta forma, los alumnos completan hasta el 3° grado del nivel primario, y se espera que continúen sus estudios en escuelas públicas o privadas luego (Linden, L., 2008).

requisitos para ser docentes en escuelas públicas, pero que fueron capacitadas para ser parte de Gyan Shala.

En este marco, se diseñó un programa de aprendizaje asistido por computadora (CAL de ahora en más, debido a sus citas en inglés) para que los alumnos trabajaran de forma independiente. Con el objetivo de que fuera accesible para familias de bajos recursos, se utiliza en computadoras usadas y antiguas que fueron donadas para el programa; que los alumnos debían utilizar por turnos (en general, se organizaban en grupos de 8 alumnos por turno).

Durante el primer año (2004-2005), se evaluó el uso del programa CAL en 23 escuelas en dos localidades indias. En cambio, durante el segundo año se implementó el programa en 37 escuelas de dos localidades diferentes y más urbanas que las anteriores.

El diseño de la intervención supuso la evaluación de dos intervenciones distintas: por un lado, se utiliza como un sustituto de la enseñanza escolar, en horario de clase; mientras que, por el otro, se usa de forma complementaria, en horario extraescolar. Así, durante el primer año, el programa CAL se utilizó en horario escolar en la segunda mitad del año, como reemplazo de horas de clase. En cambio, en el segundo año, el programa se implementó durante todo el ciclo escolar de forma extraescolar. Para usarlo fuera del horario escolar, había dos clases con el programa CAL en un día, antes y al final de las clases. Los estudiantes llegaban antes o después de la escuela dependiendo del turno de su clase. De esta manera, el programa complementó, en lugar de reemplazar, el plan de estudios básico de Gyan Shala. En ambos casos, el programa se utilizaba una hora al día, tal y como se pensó el software.

Respecto a su estrategia de identificación, se eligieron un total de 60 escuelas distribuidas en las 4 localidades participantes del estudio, 23 participantes el primer año y 37 el segundo año. Para la asignación de las escuelas, se estratificaron los resultados de los desempeños escolares promedio del año anterior de los alumnos de las escuelas, que en el próximo ciclo escolar entrarían a 2° y 3° grado.

Así, en el año 2004-2005, se realizó aleatoriamente la asignación al grupo tratamiento y al grupo control. La aleatorización estratificada resultó en un grupo de tratamiento de 11 escuelas con 392 estudiantes y un grupo de control de 12 escuelas con 387 estudiantes. En el segundo año, del mismo se asignaron 19 escuelas con 682 estudiantes al grupo de tratamiento y 18 escuelas con 695 estudiantes al grupo de control.

No hubo diferencias significativas entre las características de los grupos de tratamiento y control.

b. Diseño del software y contenido.

El programa CAL de Gyan Shala se diseñó para complementar el plan de estudios de matemáticas de cada día y para no requerir intervención alguna de ningún docente.

A su vez, el software se desarrolló de forma tal que cada computadora pudiera ser utilizada por dos alumnos al mismo tiempo, pero de forma independiente (pues la pantalla se divide en 2, y mientras uno utiliza el mouse, el otro realiza las actividades con el teclado).

Así, diseñado para usar una computadora cada dos alumnos, los ejercicios de cada día estaban relacionados, por lo menos de forma aproximada, con lo que los alumnos estaban viendo en sus clases. El programa contaba con contenido de 2° y 3° grado.

c. *Target* de la intervención

Alumnos de bajo nivel socio-económico de 2° y 3° grado de nivel primario de escuelas del programa Gyan Shala.

d. Relación software-aula

La relación con la dinámica escolar se dio de dos formas: en primer lugar, el diseño del software contempló desde su inicio el currículum que se emplea para las clases, buscando complementarlo y reforzarlo. En segundo lugar, buscando que los alumnos trabajaran contenidos similares a los que estaban viendo en sus clases mediante ejercicios asignados por el docente.

e. Evaluación del desempeño y datos recolectados

Los datos se recolectaron de tres formas: dos evaluaciones en matemática y lengua, en abril del año anterior al estudio y en abril del año del estudio, e información demográfica de los estudiantes. Los tests son los mismos que se implementaban en el programa Gyan Shala, los cuales fueron diseñados para evaluar el currículum del programa. El examen de base de línea tenía un formato diferente del examen final, aunque ambos se mantuvieron iguales en los dos años de la evaluación.

f. Resultados

Se calculó el *intention-to-treat* (ITT) con un método de diferencias en diferencias, con las escuelas como *clusters* para corregir la correlación al interior de cada escuela; sin y con controles de características de los alumnos y escuelas (*OLS*). También lo especifica con GLS (*Generalized Least Squares*), con efectos aleatorios a nivel escolar y, dentro de las escuelas, a nivel de grado; para incluir en la estimación la correlación con las características observables.

Si se analizan los dos años en conjunto, se encuentran efectos bajos y negativos (-0,018 en el *diff-en diff* con controles y -0.06 al usar *GLS*, con 0.167 y 0.154 de error estándar respectivamente), aunque no significativos. A priori, esto parecería indicar que el uso del programa CAL no fue efectivo.

Al analizar cada tipo de intervención – suplemento o complemento de las clases tradicionales-, se encuentran resultados diferentes. Así, los estimadores de los efectos del primer año de la intervención son negativos y significativos al 5%: -0.566 DS en el *diff-en diff* con controles y -0.656 DS al usar *GLS*, con errores estándar de 0.253 y 0.264 a nivel escuela respectivamente. A su vez, al analizar la distribución de los resultados por desempeños escolares, se encuentra que de los tres terciles de desempeño²¹, los resultados son sólo significativos al 5% para el primer y segundo tercil (-0.847 DS y de -0.622 DS, y sus errores estándar de 0.368 y 0.246, respectivamente) al especificarlos en el método de diferencia con controles. Por otro lado, al

²¹ Donde el primer tercil contiene a los alumnos con desempeños más bajos, y asciende en terciles de forma incremental (Linden, L., 2008).

analizar los resultados por grado, los autores encuentran resultados significativos al 5% en 2º grado, con un valor negativo de -0.741 DS en la especificación de diferencias con controles.

Durante el segundo año de intervención, en cambio, los resultados son más esperanzadores. Si bien el estimador de *OLS* no es significativo, sí se obtienen resultados significativos al emplear *GLS*. Con resultados significativos al 5% únicamente al utilizar el estimador de efectos aleatorios (*GLS*), el uso del programa CAL parece generar mejoras en los alumnos del grupo tratamiento de 0,332 DS (0.162 de error estándar, a nivel *cluster*). A su vez, al mirar la distribución de las notas, se encuentra que el programa sólo fue efectivo en los alumnos con peores desempeños en los exámenes de matemática²². Del mismo modo, al descomponer los resultados por grado, se encuentra que sólo los alumnos de 3º grado obtuvieron mejoras significativas al 5% de 0,515 DS (con un error estándar de 0.216) según el método de diferencias con controles. En cambio, los resultados de 2º grado no son significativos, lo cual podría implicar que el programa CAL no tuvo efectos en sus desempeños. Sin embargo, al analizar los efectos solo para los alumnos de bajos desempeños, los estudiantes en ambos grados obtienen mejores resultados, en 0.498 DS y 0.524 DS (y sus errores estándar de 0.193 y 0.258) para 2º y 3º grado respectivamente, significativos al 5%. Aquí, sin embargo, es preciso mencionar la considerable disminución de la cantidad de datos involucrados en el análisis²³.

Por último, el programa parece mejorar significativamente el rendimiento matemático de los estudiantes mayores y los estudiantes con el rendimiento más bajo en las escuelas.

Sobre la *attrition*, el 25% de los estudiantes de control y el 23 % de los estudiantes de tratamiento no tomaron ambas partes del examen de seguimiento, lo que sugiere que las tasas generales de deserción son muy similares. A su vez, las estimaciones realizadas para cada año demuestran que tanto la cantidad de niños como las características de los niños son similares, implicando en general a los alumnos con bajos rendimientos en la base de línea.

2.5. The Persistence of Gains in Learning from Computer Assisted Learning (CAL): Evidence from a Randomized Experiment in Rural Schools in Shaanxi Province in China (Mo, Zhang, Wang, Huang, Shi, Boswell & Rozelle, 2014)

Los autores de este estudio buscan determinar la persistencia de los efectos de un programa CAL en los resultados académicos de una población de estudiantes desatendidos en China.

a. La intervención y su estrategia de identificación

El estudio tuvo lugar en Shaanxi, una provincia China, e involucró a 5.943 estudiantes en 72 escuelas rurales en la provincia de Shaanxi, 36 escuelas del grupo de control y 36 del grupo de tratamiento. Como estas escuelas tienen alumnos que pernoctan durante la semana en la institución y, también, otros que no lo hacen; para la intervención se asignaron a tratamiento o

²² El único tercil significativo fue el primero, en ambas especificaciones. Así, obtuvo 0.398 DS en el método de diferencias con controles (0.192 de error estándar a nivel escuela), y 0.472 DS (0.194 de error estándar a nivel escuela) en el método con efectos aleatorios, ambos significativos al 5% (Linden, 2008).

²³ De una muestra en el segundo año de 631 alumnos de 2º grado y 483 de 3º grado, al analizar solo los alumnos de bajos recursos las cantidades se reducen a 146 y 174 alumnos respectivamente (Linden, 2008).

control únicamente a los alumnos que albergaban en la escuela como grupo tratamiento (*boarding students* según la investigación).

Con el objetivo de analizar los efectos según la duración de la intervención, este estudio se planteó el uso de un programa CAL en un período corto – de un semestre- y de un período largo – de año y medio-.

En una primera fase, se planteó el uso de un programa CAL de matemática que se utilizó fuera de horario escolar durante un semestre (2011) por alumnos de 3° y 5° grado. Se tomaron evaluaciones estandarizadas de matemática al antes y al finalizar el uso del programa.

Para analizar la persistencia de los resultados, se llevó adelante también una segunda fase de uso del programa, que tuvo lugar durante todo el año escolar de 2011-2012 (octubre de 2011 a junio de 2012) y se planteó como extensión del análisis de corto plazo. Incluyó a los mismos 2741 estudiantes pupilos y no pupilos de las mismas 72 escuelas rurales en la provincia de Shaanxi. Por tanto, ahora esos alumnos estaban en cuarto y sexto grado.

En ambas fases del estudio, utilizaron un software diseñado para complementar el currículum escolar en matemática, cuyas sesiones de uso se impartieron a los estudiantes bajo la supervisión de dos maestros supervisores capacitados para la intervención. Los estudiantes en el grupo de tratamiento participaron en dos sesiones de CAL de 40 minutos por semana, fuera de horario escolar. Las sesiones fueron obligatorias y facilitadas por los maestros supervisores, quienes habían sido capacitados para ello²⁴.

Durante cada sesión de CAL, los estudiantes jugaron en computadoras juegos de matemáticas diseñados para ayudarlos a repasar y practicar el material matemático básico que se enseñaba en sus clases regulares de matemáticas.

Para elegir las escuelas, se seleccionaron 4 condados dentro de Shaanxi, con el objetivo de limitarlo a zonas rurales de bajo nivel socio-económico. Luego, con el criterio de que fueran escuelas con todos los grados completos, de 1° a 6°, se eligieron 72; las cuales fueron aleatoriamente asignadas al grupo control o tratamiento. Todos los alumnos de 3° y 5° grado de las escuelas de tratamiento y control en el primer año fueron incluidos en el estudio, al igual que en el segundo año a los de 4° y 6° (es decir, a los mismos estudiantes). En total, hubo 554 estudiantes de tercer grado y 723 estudiantes de quinto grado en las escuelas de tratamiento, y 613 alumnos de 3° y 851 de 5° en las 36 escuelas del grupo de control. Los grupos no presentaron diferencias entre sí en las características de los alumnos.

b. Diseño del software y contenido.

El contenido de cada sesión, compuesto por videos instructivos y juegos, fue diseñado para ayudar a los estudiantes a alcanzar las competencias básicas en el plan de estudios nacional de matemáticas uniforme de China.

²⁴ Eran docentes de las escuelas, a quienes se le pagó un adicional para que fueran facilitadores del uso del programa. Recibieron una capacitación de dos días, y un protocolo de acción, donde se especificaba entre otras que no debían brindar asistencia a los alumnos usando el software (Mo et al, 2014)

Los maestros supervisores de CAL eran los docentes de los alumnos, a quienes se contrató por fuera de sus horas escolares para que supervisaran las sesiones de uso del software. Así, ellos debían organizar que los estudiantes se sentaran en parejas (una sola computadora era compartida por dos alumnos) y utilizaran sin inconvenientes el programa. Solo un estudiante en un momento dado tenía el control del mouse, pero a intervalos regulares se alentó a los estudiantes a turnarse para usar el mouse.

Las lecciones basadas en software fueron exactamente las mismas para todos los estudiantes del mismo grado en cada una de las escuelas de tratamiento. Con un video primero y luego con juegos que trabajaran sus contenidos, se les asignaban contenidos relacionados con la instrucción de matemática que recibían esa misma semana. El software no era, por tanto, adaptativo al nivel de conocimiento de los alumnos. De hecho, si los alumnos tenían dudas, debían discutirlos con su compañero de computadora, sin posibilidad de obtener respuesta en el momento. Al finalizar el uso del juego, el programa les decía a los estudiantes cuántas respuestas correctas e incorrectas había obtenido.

c. Target de la intervención

Alumnos de 3° y 5°, o 4° y 6° en el segundo año, de escuelas de bajo nivel socio-económico de zonas rurales.

d. Relación software-aula

La relación del software al aula se da, en primer lugar, por el diseño del contenido para reforzar el currículo nacional chino. A pesar de que eran sesiones fuera del horario escolar, el hecho de que los alumnos siguieran los contenidos de matemática que estaban viendo esa semana junto, permite suponer una relación con lo que estaba sucediendo en el aula.

Respecto a los maestros supervisores, si bien capacitados para ello, su único rol era facilitar las sesiones y asignar contenidos relacionados con la instrucción de matemática de esa semana.

e. Evaluación del desempeño y datos recolectados

Durante el experimento de un año y medio de duración, se realizó una encuesta antes del lanzamiento del programa y dos rondas de encuestas de evaluación, una al finalizar la fase I y luego otra al finalizar la fase II.

La evaluación de matemática se realizó con exámenes de matemáticas estandarizados, que relevaban los mismos contenidos, pero con distintas preguntas en cada momento de evaluación. Los contenidos eran diferentes según grado. Junto con cada instancia de evaluación, se pidió a los alumnos que completaran una encuesta autoadministrada para recolectar su información socio-económica y demográfica.

f. Resultados

Los autores estiman el *intention-to-treat (ITT)* con una regresión lineal sin y con controles por los resultados del examen inicial, características de las familias y características de los alumnos. Como en todos los estudios, toman a la escuela como sujeto de análisis para medir los errores estándar (*clusters*). Cada una de estas regresiones se repite para analizar tres extensiones del programa CAL: un semestre, un año escolar, y un año y medio (es decir, tres semestres).

Respecto a los resultados de la fase I y fase II juntas, tanto 3° como 5° grado muestran haber mejorado más que el grupo control en matemática. En la regresión con controles, los autores obtienen que los alumnos de 3° grado obtienen mejoras de 0.25 DS (0.08 de error estándar), mientras que en 5° grado de las escuelas de tratamiento mejoraron 0.26 DS (0.08 de error estándar) más que el grupo control; ambos significantes al 1%.

Sobre la fase I, y también utilizando una regresión controlada, en 3° grado los desempeños son 0.18 DS (0.08 de error estándar) mejores que en el grupo control, con un nivel de significancia al 5%. En quinto grado, en cambio, los resultados son no significativos. Al analizar la fase II por separado, en 3° grado el grupo tratamiento mantiene sus mejoras respecto al grupo control, pero con un efecto no significativo; mientras que se invierte la situación de 5° grado, pues obtiene efectos positivos de 0,15 DS respecto al grupo control, significativos al 10% y con 0.08 de error estándar.

Al analizar la heterogeneidad de estos resultados, los autores no encuentran en este estudio que haya diferencias en los efectos que el programa CAL tuvo según los desempeños de los alumnos en el examen inicial.

RESULTADOS

Las investigaciones analizadas en esta revisión incluyen gran parte de las evaluaciones experimentales aleatorizadas sobre el uso de programas CAL en el aprendizaje de matemática en nivel primario o medio de países en desarrollo²⁵. Sin embargo, no busca ser una revisión exhaustiva, sino señalar posibles direcciones que la evidencia sugiere para países en desarrollo. Para ello, es preciso señalar que no existe al día de hoy una revisión que se concentre en el área de matemática, a pesar de que parece ser el área donde los programas CAL podrían ser más eficientes (Escueta et al, 2017).

En la tabla 1, se muestra una comparación de las características de cada evaluación y sus resultados.

²⁵ Como se mencionó antes, la única de las evaluaciones experimentales llevadas adelante en países en desarrollo sobre matemática excluida fue la de Lai et al. (2015), en tanto con metodología y resultados similares a (Mo et al, 2014), su muestra era más pequeña y un público particular de China (*migrant schools* en inglés, que son escuelas de familias de bajo nivel socio-económico que se trasladan de zonas rurales a zonas urbanas).

Estudio	País	Uso del software	Software adaptativo	Método	Efectos en matemática	Heterogeneidad	Frecuencia de uso	Duración	Destinatarios (grado de alumnos)	Observaciones
Banerjee, Cole, Duflo & Linden, 2007	India	Complemento y Sustituto	Si	OLS con controles	Año 1: 0.35 DS Año 2: 0.48 DS	Tercil 1: 0.425 DS Tercil 2: 0.316 DS Tercil 3: 0.266 DS	2 sesiones semanales de una hora cada una	Todo el año escolar	4° grado de nivel primario	Intervención de tutoría en mismas escuelas. El análisis por terciles contempló solo una cohorte del estudio (muestra más pequeña).
Muradliharan, Singh & Ganimian, 2017	India	Complemento	Si	OLS con efectos fijos	0.36 DS***	No signif.	2 sesiones semanales de 90 minutos cada una	4,5 meses	6° a 9° grado de nivel medio, de bajo nivel socio-económico	Intervención de <i>blended learning</i>
Araya, Ortiz, Bottan & Cristia, 2019	Chile	Complemento y Sustituto	No	OLS con efectos fijos y controles	0.27 DS***	No signif.	2 sesiones semanales de 90 minutos cada una	7 meses	4° grado de primaria con bajos rendimientos en matemática	Análisis con los exámenes estandarizados nacionales. El realizado con los exámenes desarrollados para el estudio dan resultados considerablemente más bajos.
Linden, 2008	India	Sustituto	No	OLS con controles	-0.566 DS**	Tercil 1: -0.847 DS** Tercil 2: -0.622 DS***	1 hora diaria	1 semestre	2° y 3° grado de primaria de bajo nivel socio-económico	Al desagregar por grados, sólo obtiene resultados significativos 2° grado
		Complemento		OLS con controles	No signif.	Tercil 1: 0.472 DS**	1 hora diaria	Todo el año escolar	2° y 3° grado de primaria de bajo nivel socio-económico	Al desagregar por grados, sólo obtiene resultados significativos 3° grado

Mo, Zhang, Wang, Huang, Shi, Boswell & Rozelle, 2014	China	Complemento	No	OLS con controles	3° grado: 0.18 DS**	No signif.	2 sesiones semanales de 40 minutos cada una	1 semestre	3° y 5° grado de primaria de escuelas rurales	-
					6° grado: 0.15 DS*	No signif.	2 sesiones semanales de 40 minutos cada una	Todo el año escolar	4° y 6° grado de primaria de escuelas rurales de bajo nivel socio-económico	Extensión del estudio corto
					3° grado: 0.25 DS***, 5° grado: 0.26 DS ***	No signif.	2 sesiones semanales de 40 minutos cada una	Año escolar y medio	3°/4° y 5°/6° grado de escuelas primarias rurales de bajo nivel socio-económico	Fases I y II juntas

NOTA: en todos los casos se mencionan los *ITT* de las intervenciones. Sin embargo, como provienen de intervenciones distintas en términos de duración temporal, especificación metodológica empleada (controles), tamaños y composición de muestro e incluso en algunos casos de intervenciones radicalmente distintas (por ejemplo, de *blended learning*), esta comparación debería tomarse de forma cautelosa. En *heterogeneidad*, el tercil 1 corresponde a los alumnos con desempeños más bajos, el 2 al de desempeños medios y el 3 al de desempeños más altos. Se detallan los niveles de significancia declarados por los autores con * para el 10 %, ** 5% y *** 1%.

A partir de la literatura revisada, es posible plantearse ciertos interrogantes que, si bien podrían indicar patrones, aún falta responder.

Uno de los principales interrogantes es si la eficiencia del programa CAL en términos de aprendizaje se encuentra relacionada con la personalización y adaptabilidad de los contenidos. Es decir, si es más eficiente un software cuyo contenido y complejidad se ajusta acorde al nivel y progreso de cada alumno; o si lo es un software donde todos los alumnos trabajan los mismos contenidos sin distinguir sus habilidades, nivel de desempeño y puntos débiles. Por un lado, vemos que tan solo dos de los programas analizados son adaptativos, mientras que el resto presentan a los alumnos juegos propios del grado correspondiente. Una posible explicación de porqué estos programas favorecerían más a los alumnos es que, al adecuarse al nivel de cada uno, los alumnos pueden aprender desde la base que poseen, y no la esperada por el sistema educativo (Banerjee et al, 2017). Ahora bien, el resto de los estudios analizados aquí no son adaptativos y algunos poseen aun así resultados positivos. Esto podría explicarse por dos motivos: (i) en el caso de los alumnos que ya comprendieron los contenidos enseñados, sería lógico esperar que el uso de tecnología no mejore su desempeño (Linden, 2008), o lo que es lo mismo, estos programas ayudan a reforzar contenidos de clase no comprendidos enteramente o (ii) bien por otros efectos relacionados con el uso de la tecnología, ya sea por el efecto de pares cuando utilizan dos alumnos una computadora al mismo tiempo (Mo et al, 2014), por la motivación intrínseca a algo novedoso, u otros factores sobre los cuales debería profundizarse. Más allá de lo poco que la evidencia puede señalar al respecto, es posible observar que las dos intervenciones con software adaptativo son las que obtuvieron efectos más altos, sugiriendo que las intervenciones con programas CAL adaptativos podrían resultar en mejoras en el desempeño más altas.

Una pregunta relacionada con la anterior es, por tanto, si los programas CAL contribuyen a que los alumnos con peor desempeño mejoren más que el resto, o si posee los mismos efectos en todos los alumnos. De las investigaciones revisadas, solo dos de ellas encuentran efectos mayores en los alumnos con desempeños más bajos, al desagregarlos del resto. Además, también se observa que uno de los casos analizados encuentra efectos negativos en el desempeño los dos terciles más altos (Linden, 2008), es decir, que los alumnos con mejores desempeños empeoraron en relación al grupo control. Esto podría ser porque los docentes suelen enseñar según el currículum, sin enfocarse necesariamente en ayudar a los alumnos que están rezagados, volviendo la enseñanza inefectiva para ellos (Banerjee et al, 2007). De hecho, en el caso de softwares adaptativos y a diferencia de un docente, estos pueden adecuar el contenido al nivel de cada alumno sin costos marginales (Muradliharan et al, 2017). Lo sorprendente aquí es que una de las intervenciones implicó el uso de un software no adaptativo; lo cual nos lleva a preguntarnos si el nivel de desempeño de los estudiantes podría explicar en qué contextos es más efectivo un software adaptativo y en cuáles es más efectivo el que no lo es. A priori, la evidencia podría sugerir que, si los alumnos poseen rezagos grandes que condicionan los aprendizajes curriculares esperados para su grado, es posible que el software adaptativo sea más eficiente que uno que entrena habilidades o enseña conocimientos en dicho nivel. En cambio, allí donde los alumnos simplemente no comprenden a la perfección los contenidos -pero tengan una comprensión de ellos-, es esperable que un software que los refuerce mejore los desempeños (Linden, 2008).

Otro gran interrogante que surge de la literatura revisada es si es más conveniente utilizar programas CAL como reemplazo o complemento del docente. En este sentido, se entiende que un programa es un sustituto del aula cuando se usa en horario escolar y, por tanto, reemplaza horas que se destinarían a la clase de matemática. Por otro lado, se entiende complemento a aquellas intervenciones que, independientemente de ser facilitadas o no por el docente en cuestión, buscan reforzar contenidos aprendidos en la escuela a través de horas adicionales de enseñanza de matemática. Aquí la investigación de Linden (2008) podría ser la más informativa, ya que encuentra que la forma de implementarlo afecta la eficiencia del programa: al utilizarlo como sustituto de los recursos existentes, el programa obtiene efectos negativos y significativos; mientras que cuando se utiliza como complemento a la situación existente, los resultados demuestran que hay mejoras de 0.28 desvíos estándar (Linden, 2008). Sin embargo, las dos intervenciones analizadas se implementaron con duraciones distintas, por lo que esta comparación debe realizarse con precaución. En cuanto a la evidencia existente en países desarrollados, Räsänen et al (2019) señala que en los casos donde los programas se utilizan de forma complementaria suelen ser más efectivos, aunque no discrimina entre intervenciones en países en desarrollo y desarrollados. Como la mayoría de los estudios revisados aquí utilizan un mix de ambas modalidades, pues tan sólo la investigación de Linden implementa el programa CAL como sustituto, cabría profundizar en este análisis.

A su vez, la pregunta sobre los efectos de largo plazo se encuentra hoy inconclusa con la evidencia disponible. De los estudios analizados, tan solo Banerjee et al (2007) abordan este punto. Encuentra resultados significativos y positivos, pero mucho más bajos que los de corto plazo (0.09 desvíos estándar en comparación con 0.35 DS y 0.47 DS del primer y segundo año de la intervención, correspondientemente). Sin embargo, es preciso señalar que aquí el diseño de la intervención que es, por lo menos, complejo, implicó una muestra notoriamente más chica para analizar el segundo plazo y que algunos alumnos de bajo desempeño escolar fueron participantes de otra intervención, lo cual podría generar efectos confundibles. Son varios los motivos por los que pocos estudios ahondan en este aspecto. Uno de ellos es que es complejo y caro realizar el seguimiento durante tanto tiempo en evaluaciones de impacto (Banerjee et al, 2017). Además, dada la velocidad de los avances tecnológicos, es posible que la intervención se vuelva obsoleta a lo largo de la evaluación, motivo por el cual los costos se incrementan aún más.

Por otro lado, un gran interrogante inconcluso aquí es cuál es la frecuencia y duración óptimas para este tipo de intervenciones. Nuevamente, solo uno de los estudios busca responder esta pregunta y trata, puntualmente, sobre la duración de la intervención. Mo et al (2014) obtiene resultados ampliamente más grandes al analizar la misma intervención durante un semestre que durante un año escolar y medio. Sin embargo, es preciso aclarar que los efectos de la intervención medidos por fase (la primera, de un semestre; la segunda, de un año escolar) no son contundentes: analizando los resultados de dos grados, en una fase los resultados solo son significativos para 3° grado, mientras que en la segunda fase se invierte la significatividad de los resultados (es decir, solo en 5° grado son significativos). Así, no sólo se encuentran diferencias en los grados en los que se obtienen mejoras significativas, sino también en la magnitud de estas mejoras (ya que las fases por separado obtienen efectos entre 7 y 9 DS inferiores. Respecto a la frecuencia, no es posible establecer un patrón claro, puesto que solo pocos de los estudios analizados mencionan el uso efectivo de la plataforma. Sin embargo, al mirar el tratamiento

asignado (y no el efectivo) es posible ver que las intervenciones se brindan, en promedio, dos veces por semana. De hecho, la única intervención que incrementa su frecuencia es la de Linden (2008), que obtiene coeficientes negativos y/o no significativos, salvo en su análisis de GLS.

Otro interrogante válido es si hay una población objetivo donde este tipo de intervenciones presente mejores resultados. De lo analizado en esta revisión, tan solo un estudio (Murahlidan et al, 2017) trabaja en el nivel medio – aunque con alumnos cuyo desfasaje escolar los ubicaría en el nivel primario-, mientras que el resto son todos realizados en escuelas primarias. Esto puede deberse, como se ha mencionado previamente, al hecho de que el nivel primario suele ser el nivel con mayor cobertura a nivel global y que, por tanto, este nivel permite el alcance a poblaciones más vulnerables. Por otro lado, una conclusión que se extrae de los casos revisados es que en estos contextos de bajos recursos socio-económicos las intervenciones con programas CAL son efectivas. Esto, a su vez, coincide con la evidencia de países desarrollados (Escueta et al, 2017). Del mismo modo, se observa que la mayor parte de las intervenciones involucran niños ya escolarizados y que por lo menos han completado el primer grado del nivel primario. Esto podría deberse, por un lado, a la dificultad de evaluar a niños que, no sólo no saben leer y escribir ni conocen los números, sino que aún están acostumbrándose a la situación de escolaridad en general. Del mismo modo, podría adjudicarse a que cuando los niños son un poco más grandes, poseen más capacidades para encontrarse en situaciones de aprendizaje computarizado sin contar asistencia de un adulto. Finalmente, otra hipótesis podría ser que es más fácil ampliar el rango de contenidos a algo más complejo (grados superiores), que a niveles más simples. Esto se debe no sólo al tipo de soporte que los niños necesitan en cada etapa, sino también a que en términos de evaluación permite incorporar más grados dentro de una misma institución – es decir, de nivel primario-.

DISCUSIÓN

A partir de lo anterior, vemos que la forma en que estas intervenciones se implementan y se relacionan con su contexto importa tanto como el diseño del programa. La inserción de los programas CAL parece requerir especial atención en ciertos puntos. En primer lugar, incorporar recursos, softwares en este caso, sólo funcionará si éstos responden a necesidades insatisfechas de la escuela (Banerjee et al, 2007). Los programas CAL pueden responder diferentes necesidades, como la escasez de docentes, la baja calidad en la enseñanza, el escaso tiempo de clase, el rezago en el desempeño de los alumnos. Sin embargo, pensar cuál de ellos y de qué forma se busca responderlo, parece condición para que el programa sea efectivo en el aprendizaje. Así, por ejemplo, en comunidades educativas con escasez de docentes, sería esperable que el software generara efectos positivos tanto si se usa dentro como fuera del aula, pues el escenario opuesto sería la ausencia de un docente. Es posible pensar esta reflexión también en casos menos extremos. Por ejemplo, allí donde los alumnos presentan grandes rezagos escolares, un software adaptativo (y una evaluación acorde a los contenidos) probablemente permita un mayor aprendizaje por parte de los alumnos, ya que les permitirá trabajar contenidos que difícilmente un docente pueda abordar en clase si hay una gran heterogeneidad de conocimientos de sus alumnos.

Por otro lado, es preciso considerar cuidadosamente la forma en que los nuevos recursos interactuarán con los existentes (Linden, 2008). Así, si el uso será dentro del aula, si se implementará con la ayuda docente, o si por el contrario su uso será autónomo e independiente del mismo no son factores menores. De este modo, vemos que no sólo se trata de necesidades que cubrir, sino también del contexto en el que se dan estas necesidades. Así, por ejemplo, en zonas rurales aisladas, se vuelve necesario no sólo que el software funcione sin internet, sino también que sea el mismo docente el que contribuya en su intervención -pues en general la escuela no posee otro personal-. En cambio, en escuelas de zonas urbanas donde las escuelas presentan grandes matrículas, el uso de un programa CAL autónomo -que permita a los alumnos seguir su propio paso, y no el de un grupo grande de niños- podría presentar efectos más altos.

En tercer lugar, la adecuación del contenido resulta sumamente importante. En todas las evaluaciones analizadas, el software había sido especialmente diseñado para responder en términos del currículum nacional. Independientemente de la adaptabilidad del software, es posible que esto se relacione más con la forma de evaluación que con el proceso de enseñanza. Es decir, que la intervención – programas CAL- fuera acorde a los contenidos que se buscaba mejorar y coherente con lo relevado en la evaluación.

En cuarto lugar, el hecho de que a excepción de una intervención todos los estudios encuentren resultados positivos sugiere que, efectivamente, este tipo de tecnología resulta más prometedora en países en desarrollo, donde hay grandes inequidades y necesidades insatisfechas, que en países desarrollados (Escueta et al, 2017). Respecto a qué intervención produce mejores efectos, aquí la gran heterogeneidad de intervenciones dentro de los pocos estudios analizados no permite hacer afirmaciones contundentes. Sin embargo, de algo no hay dudas: el uso de programas CAL es una gran ventana de oportunidad para mejorar la educación en aquellos países que, con pocos recursos, deben enfrentar grandes brechas socio-económicas y educativas.

CONCLUSIONES

A lo largo de esta investigación, se revisaron cinco estudios de gran relevancia en la literatura científica sobre uso de programas CAL de matemática en contextos educativos de países en desarrollo. Sin la pretensión de sacar conclusiones generalizadas, los resultados aquí encontrados enfatizan la importancia del diseño de una intervención – en qué medida es necesaria y factible en un contexto dado- y cómo se adapta a la necesidad de los alumnos y su heterogeneidad. En sí, la tecnología es tan solo un medio de aprendizaje que debe ajustarse a un fin o, de lo contrario, caer en el *solucionismo tecnológico* puede llegar a ser perjudicial para el aprendizaje de los alumnos.

REFERENCIAS

Araya, R., Arias Ortiz, E., Bottan, N., & Cristia, J. (2019). ¿Funciona la gamificación en la educación?: evidencia experimental de Chile. *Banco Interamericano de Desarrollo*. Documento de trabajo; 982. <http://dx.doi.org/10.18235/0001777>

Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007). Remedying Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics*, 122: 1235–1264. <https://www.nber.org/papers/w11904.pdf>

Barrera-Osorio, Felipe & Linden, Leigh. (2009). The Use and Misuse of Computers in Education: Evidence from a Randomized Experiment in Colombia. *The World Bank*, Policy Research Working Paper Series. <http://documents.worldbank.org/curated/en/346301468022433230/pdf/WPS4836.pdf>

Carrillo, P., Onofa, M., & Ponce, J. (2011): Information Technology and Student Achievement: Evidence from a Randomized Experiment in Ecuador, IDB Working Paper Series, No. IDB-WP-223, Inter-American Development Bank (IDB), Washington, DC

Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88e113.

Cristia, J. P., Ibararán, P., Cueto, S., Santiago, A. & Severin, E. (2012). Tecnología y desarrollo en la niñez: Evidencia del programa Una Laptop por Niño. IDB working paper series, 304.

Escueta, M., Quan, V., Nickow, A. J., & Oreopoulos, P. (2017). *Education technology: an evidence-based review*. NBER Working Paper No. 23744, National Bureau of Economic Research.

Glewee, P., & Muradlihan, K. (2015). Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications. *Handbook of the Economics of Education*. <https://doi.org/10.1016/B978-0-444-63459-7.00010-5>

Kulik, J. A. (2003). Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say (SRI Project No. P10446.001). Arlington, VA: SRI International. https://www.ic.unicamp.br/~wainer/cursos/2s2004/impactos2004/Kulik_ITinK-12_Main_Report.pdf

Mo, D., Zhang, L., Luo, R., Qu, Q., Huang, W., Wang, J., Qiao, Y., Boswell, M., & Rozelle, S. (2014). Integrating Computer-Assisted Learning into a Regular Curriculum: Evidence from a Randomized Experiment in Rural Schools in Shaanxi. *Journal of Development Effectiveness*, 6: 300–323. <https://doi.org/10.1080/19439342.2014.911770>

Muradlihan, K., Singh, A., & Ganimian, A. (2017). Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India. *NBER Working Paper* No. 22923, December 2016, Revised July 2017. DOI: 10.1257/aer.20171112

Murphy, R., Penuel, W., Means, B., Korbak, C., Whaley, A., & Allen, J. (2002). E-DESK: A review of recent evidence on discrete educational software (SRI International Report). Menlo

Park, CA: SRI International.
https://archive.sri.com/sites/default/files/publications/imports/Task3_FinalReport3.pdf

Linden, L. (2008). Complement or substitute? The effect of technology on student achievement in India. Working Paper. Columbia University Department of Economics.
<http://documents.worldbank.org/curated/en/804371468034237060/pdf/448630NWP0Box31er17010Gyan0Shala111.pdf>

Laia, L., Zhang, H., & Rozelle, S. (2015). Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment immigrant schools in Beijing. *Economics of Education Review*. <https://doi.org/10.1016/j.econedurev.2015.03.005>

Räsänen, P., Laurillard, D., Käser, T., & von Aster, M. (2019). Perspectives to Technology-Enhanced Learning and Teaching in Mathematical Learning Difficulties. In: *International Handbook of Mathematical Learning Difficulties* (pp. 733-754). Springer, Cham.